# Google’s opt-out protections fail to stop publisher content use in AI, says DeepMind VP



In a recent testimony during ongoing antitrust proceedings, Eli Collins, Vice President of Product at Google DeepMind, unveiled crucial insights regarding the utilisation of publisher content in Google’s search products. The case, which seeks to address accusations of Google holding a monopoly over the search and artificial intelligence (AI) sectors, has brought to light complex dynamics affecting content creators.

Collins explained that while Google DeepMind's AI models respect the preferences of publishers who have opted out of AI training, Google’s search products are not bound by the same rules. Specifically, a significant portion of the data used to train these AI models—amounting to 80 billion tokens—originated from content belonging to publishers who had chosen to opt out. However, he clarified that once a publisher actively opts out, their content is excluded from DeepMind's AI models.

The nuances of this situation became evident with the discussion surrounding Google’s Gemini AI model. Collins acknowledged that content might still be incorporated into Google’s Search functionalities, provided it aligns with the operation of features such as AI Overviews and the newly launched AI Mode. This indicates that traditional opt-out measures may fall short in preventing Google from using publisher content, raising substantial concerns about the level of control publishers maintain over their material.

Further complicating matters, Google's updated privacy policy from June 2023 broadened the scope of data classified as publicly available, allowing the tech giant to use any website data that does not require a paywall or mandatory sign-up for training its language models. This position was clarified by a Google spokesperson, who noted that the only way for publishers to prevent their data from being utilised in Search AI is to opt out of being indexed altogether. This is accomplished by disabling the robots.txt standard, which governs how Google indexes web content.

However, this approach presents a daunting dilemma for publishers. Should they choose to opt out of indexing, their website pages would vanish from Google search results, effectively limiting their online visibility while leaving their content vulnerable to AI training. As a result, publishers find themselves grappling with the implications of an evolving digital landscape increasingly dominated by AI technologies.

The antitrust case itself, pursued by the Department of Justice, aims to scrutinise Google's market position and its practices. US District Judge Amit Mehta is being urged to compel Google to divest its Chrome browser and to provide transparency regarding the data informing its search results. Notably, there have been no calls to impose similar obligations regarding the use of data for Google's AI products, setting the stage for a potentially uneven regulatory landscape.

As these legal proceedings unfold, the implications for digital content creation and the balance of power between tech giants and publishers remain critical focal points of discussion. The outcomes may shape not only the future of Google but also the broader interactions between technology and content creation in an increasingly AI-driven environment.

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.reuters.com/legal/googles-ai-previews-erode-internet-edtech-company-says-lawsuit-2025-02-24/> - This article discusses a lawsuit filed by Chegg, an educational technology company, against Google, alleging that Google's AI-generated content is undermining the digital publishing industry by reducing demand for original work and driving users away from publishers' websites. This supports the claim that publishers are concerned about the impact of AI on their content and revenue.
2. <https://www.searchenginejournal.com/google-updates-privacy-policy-to-collect-public-data-for-ai-training/490715/> - This article reports on Google's update to its privacy policy, allowing the company to collect and analyze publicly available information to train its AI models. This corroborates the claim that Google's updated privacy policy from June 2023 broadened the scope of data classified as publicly available for AI training purposes.
3. <https://www.congress.gov/event/118th-congress/senate-event/LC73543/text> - This document provides the transcript of a Senate hearing titled 'Oversight of A.I.: The Future of Journalism,' where discussions centered on the use of publisher content for AI training and the need for licensing and compensation. This supports the claim that there are ongoing discussions about the use of publisher content in AI and the need for proper licensing and compensation.
4. <https://www.wsls.com/business/2024/09/26/google-expert-at-antitrust-trial-says-government-underestimates-competition-for-online-ad-dollars/> - This article reports on testimony from an expert hired by Google during an antitrust trial, who argued that the government's claims about Google's monopoly over advertising technology are improperly focused on a narrow market definition. This supports the claim that the Department of Justice is scrutinizing Google's market position and practices.
5. <https://www.wsls.com/business/2024/09/19/at-google-antitrust-trial-documents-say-one-thing-the-tech-giants-witnesses-say-different/> - This article discusses discrepancies between documents presented in the Google antitrust trial and the testimonies of Google's witnesses, highlighting the complexities of the case. This supports the claim that the antitrust case involves complex dynamics affecting content creators and Google's practices.
6. <https://archive.org/details/CSPAN3_20240916_163700_Discussion_on_Emerging_AI_Technologies> - This video archive features a discussion on emerging AI technologies, including the use of publisher content for AI training and the need for licensing and compensation. This supports the claim that there are ongoing discussions about the use of publisher content in AI and the need for proper licensing and compensation.
7. <https://news.google.com/rss/articles/CBMimwFBVV95cUxNajNFS0R6UVd2bm1GRE9rel9wN0hfZTZSRWtsMVRtUVBfdGs4YmVlakJmbTJ4QXZNQU1BTnNGbHZMR0pCdWNUM251S0FOUkV0bmxhY3Nmd1lnQ2dlai1ock94c3JWa2JPX2Y4RXlRTnltbDR3MXVhVVVpVExyeUd5VVVaQXVsU3dOQk5OWGtpcU1YS2o2UEJGSXVFMA?oc=5&hl=en-US&gl=US&ceid=US:en> - Please view link - unable to able to access data